CLAIMS

What is claimed is:

1. A method for evaluating similarity among a plurality of data structures, comprising:

analyzing each structure of said plurality of data structures to generate at least one substructure;

matching said at least one substructure to a database having a plurality of entries to obtain at least one matching entry; and

generating a match value using a relative entropy value corresponding to said at least one matching entry.

2. The method according to claim 1, further comprising:

creating said plurality of entries in said database; and

processing said plurality of entries in said database.

3. The method according to claim 2, wherein said creating further comprises creating said plurality of entries using a tool having a graphical user interface and exporting said plurality of entries to said database.

4. The method according to claim 2, wherein said processing further comprises:

verifying said plurality of entries for validity; and

calculating said relative entropy value corresponding to each entry of said plurality of entries.

5. The method according to claim 4, wherein said processing further comprises storing said each entry of said plurality of entries together with said corresponding relative entropy value in a compressed format.

6. The method according to claim 1, further comprising extracting from a lexicon database having a plurality of elements each element associated to said each structure, assigning at least one code of said each element to said each structure, and retrieving said at least one code during matching to obtain said at least one matching entry.

7. The method according to claim 6, further comprising reading lexical probability files and assigning a probability value to said each element of said plurality of elements in said lexicon database.

8. The method according to claim 1, wherein each structure of said plurality of data structures is a representation of a linguistic expression.

9. The method according to claim 4, wherein said database is a thesaurus hierarchy including a root entry, said plurality of entries depending from said root entry.

10. The method according to claim 9, wherein said relative entropy value corresponding to said each entry of said plurality of entries is calculated based on an entropy value of said each entry and an entropy value of said root entry.

11. The method according to claim 6, wherein said each element in said lexicon database is a word.

12. A method for evaluating similarity among a plurality of data structures comprising:

creating a plurality of entries in a database; and

calculating a relative entropy value corresponding to each entry of said plurality of entries.

13. The method according to claim 12, further comprising storing said each entry of said plurality of entries together with said corresponding relative entropy value in a compressed format.

14. The method according to claim 12, further comprising:

creating said plurality of entries using a tool having a graphical user interface; and

exporting said plurality of entries to said database.

15. The method according to claim 12 further comprising:

analyzing each structure of said plurality of data structures to generate at least one substructure;

matching said at least one substructure of said each structure to said database to obtain at least one matching entry; and

generating a match value using said relative entropy value corresponding to said at least one matching entry.

16. The method according to claim 15, further comprising:

verifying said plurality of entries for validity;

extracting from a lexicon database having a plurality of elements each element associated to said each structure;

reading lexical probability files;

assigning a probability value to said each element of said plurality of elements in said lexicon database;

assigning at least one code of said each element to said each structure; and

retrieving and matching said at least one code to said database to obtain said at least one matching entry.

17. The method according to claim 16, wherein said each structure of said plurality of data structures is a representation of a linguistic expression.

18. The method according to claim 12, wherein said database is a thesaurus hierarchy including a root entry, said plurality of entries depending from said root entry.

19. The method according to claim 18, wherein said relative entropy value corresponding to said each entry of said plurality of entries is calculated based on an entropy value for said each entry and an entropy value for said root entry.

20. The method according to claim 15, wherein said each element in said lexicon database is a word.

21. A computer readable medium containing executable instructions which, when executed in a processing system, cause the system to perform a method for evaluating similarity among a plurality of data structures, the method comprising:

analyzing each structure of said plurality of data structures to generate at least one substructure;

matching said at least one substructure to a database having a plurality of entries to obtain at least one matching entry; and

generating a match value using a relative entropy value corresponding to said at least one matching entry.

22. The computer readable medium according to claim 21, wherein the method further comprises:

creating said plurality of entries in said database; and

processing said plurality of entries in said database.

23. The computer readable medium according to claim 22, wherein said creating further comprises creating said plurality of entries using a tool having a graphical user interface and exporting said plurality of entries to said database.

24. The computer readable medium according to claim 22, wherein said processing further comprises:

verifying said plurality of entries for validity; and

calculating said relative entropy value corresponding to each entry of said plurality of entries.

25. The computer readable medium according to claim 24, wherein said processing further comprises storing said each entry of said plurality of entries together with said corresponding relative entropy value in a compressed format.

26. The computer readable medium according to claim 21, further comprising extracting from a lexicon database having a plurality of elements each element associated to said each structure, assigning at least one code of said each element to said each structure, and retrieving said at least one code during matching to obtain said at least one matching entry.

27. The computer readable medium according to claim 26, further comprising reading lexical probability files and assigning a probability value to said each element of said plurality of elements in said lexicon database.

28. The computer readable medium according to claim 21, wherein each structure of said plurality of data structures is a representation of a linguistic expression.

29. The computer readable medium according to claim 24, wherein said database is a thesaurus hierarchy including a root entry, said plurality of entries depending from said root entry.

30. The computer readable medium according to claim 29, wherein said relative entropy value corresponding to said each entry of said plurality of entries is calculated based on an entropy value of said each entry and an entropy value of said root entry.

31. The computer readable medium according to claim 26, wherein said each element in said lexicon database is a word.

32. A computer readable medium containing executable instructions which, when executed in a processing system, cause the system to perform a method for evaluating similarity among a plurality of data structures, the method comprising:

creating a plurality of entries in a database; and

calculating a relative entropy value corresponding to each entry of said plurality of entries.

33. The computer readable medium according to claim 32, further comprising storing said each entry of said plurality of entries together with said corresponding relative entropy value in a compressed format.

34. The computer readable medium according to claim 32, further comprising:

creating said plurality of entries using a tool having a graphical user interface; and

exporting said plurality of entries to said database.

35. The computer readable medium according to claim 32 further comprising:

analyzing each structure of said plurality of data structures to generate at least one substructure;

matching said at least one substructure of said each structure to said database to obtain at least one matching entry; and

generating a match value using said relative entropy value corresponding to said at least one matching entry.

36. The computer readable medium according to claim 35, further comprising:

verifying said plurality of entries for validity;

extracting from a lexicon database having a plurality of elements each element associated to said each structure;

reading lexical probability files;

assigning a probability value to said each element of said plurality of elements in said lexicon database;

assigning at least one code of said each element to said each structure; and

retrieving and matching said at least one code to said database to obtain said at least one matching entry.

37. The computer readable medium according to claim 36, wherein said each structure of said plurality of data structures is a representation of a linguistic expression.

38. The computer readable medium according to claim 32, wherein said database is a thesaurus hierarchy including a root entry, said plurality of entries depending from said root entry.

39. The computer readable medium according to claim 38, wherein said relative entropy value corresponding to each entry of said plurality of

entries is calculated based on an entropy value for said each entry and an entropy value for said root entry.

40. The computer readable medium according to claim 35, wherein said each element in said lexicon database is a word.

41. An article of manufacture comprising a program storage medium readable by a computer and tangibly embodying at least one program of instructions executable by said computer to perform method steps for evaluating similarity among a plurality of data structures, said method comprising:

analyzing each structure of said plurality of data structures to generate at least one substructure;

matching said at least one substructure to a database having a plurality of entries to obtain at least one matching entry; and

generating a match value using a relative entropy value corresponding to said at least one matching entry.

42. The article of manufacture according to claim 41, wherein the method further comprises:

creating said plurality of entries in said database; and

processing said plurality of entries in said database.

43. The article of manufacture according to claim 42, wherein said creating further comprises creating said plurality of entries using a tool having a graphical user interface and exporting said plurality of entries to said database.

44. The article of manufacture according to claim 42, wherein said processing further comprises:

verifying said plurality of entries for validity; and

calculating said relative entropy value corresponding to each entry of said plurality of entries.

45. The article of manufacture according to claim 44, wherein said processing further comprises storing each entry of said plurality of entries together with said corresponding relative entropy value in a compressed format.

46. The article of manufacture according to claim 41, wherein the method further comprises:

extracting from a lexicon database having a plurality of elements each element associated to said each structure;

assigning at least one code of said each element to said each structure; and

retrieving said at least one during matching to obtain said at least one matching entry.

47. The article of manufacture according to claim 46, wherein the method further comprises reading lexical probability files and assigning a probability value to said each element of said plurality of elements in said lexicon database.

48. The article of manufacture according to claim 41, wherein each structure of said plurality of data structures is a representation of a linguistic expression.

49. The article of manufacture according to claim 44, wherein said database is a thesaurus hierarchy including a root entry, said plurality of entries depending from said root entry.

50. The article of manufacture according to claim 49, wherein said relative entropy value corresponding to said each entry of said plurality of entries is calculated based on an entropy value of said each entry and an entropy value of said root entry.

51. The article of manufacture according to claim 46, wherein said each element in said lexicon database is a word.

52. An article of manufacture comprising a program storage medium readable by a computer and tangibly embodying at least one program of instructions executable by said computer to perform method steps for evaluating similarity among a plurality of data structures, said method comprising:

creating a plurality of entries in a database; and

calculating a relative entropy value corresponding to each entry of said plurality of entries.

53. The article of manufacture according to claim 52, wherein the method further comprises storing said each entry of said plurality of entries

together with said corresponding relative entropy value in a compressed format.

54. The article of manufacture according to claim 52, wherein the method further comprises:

creating said plurality of entries using a tool having a graphical user interface; and

exporting said plurality of entries to said database.

55. The article of manufacture according to claim 52, wherein the method further comprises:

analyzing each structure of said plurality of data structures to generate at least one substructure;

matching said at least one substructure of said each structure to said database to obtain at least one matching entry; and

generating a match value using said relative entropy value corresponding to said at least one matching entry.

56. The article of manufacture according to claim 55, wherein the method further comprises:

verifying said plurality of entries for validity;

extracting from a lexicon database having a plurality of elements each element associated to said each structure;

reading lexical probability files;

assigning a probability value to said each element of said plurality of elements in said lexicon database;

assigning at least one code of said each element to said each structure; and

retrieving and matching said at least one code to said database to obtain said at least one matching entry.

57. The article of manufacture according to claim 56, wherein said structure of said plurality of data structures is a representation of an linguistic expression.

58. The article of manufacture according to claim 52, wherein said database is a thesaurus hierarchy including a root entry, said plurality of entries depending from said root entry.

59. The article of manufacture according to a claim 58, wherein said relative entropy value corresponding to said each entry of said plurality of entries is calculated based on an entropy value for said each entry and an entropy value for said root entry.

60. The article of manufacture according to claim 55, wherein said each element in said lexicon database is a word.

61. A system for evaluating similarity among a plurality of data structures, comprising:

means for analyzing each structure of said plurality of data structures to generate at least one substructure;

means for matching said at least one substructure to a database having a plurality of entries to obtain at least one matching entry; and

means for generating a match value using a relative entropy value corresponding to said at least one matching entry.

62. The system according to claim 61, further comprising:

means for creating said plurality of entries in said database; and

means for processing said plurality of entries in said database.

63. The system according to claim 62, wherein said creating means further comprises means for creating said plurality of entries using a tool having a graphical user interface and exporting said plurality of entries to said database.

64. The system according to claim 62, wherein said processing means further comprises:

means for verifying said plurality of entries for validity; and

means for calculating said relative entropy value corresponding to each entry of said plurality of entries.

65. The system according to claim 64, wherein said processing means further comprises means for storing said each entry of said plurality of entries together with said corresponding relative entropy value in a compressed format.

66. The system according to claim 61, further comprising:

means for extracting from a lexicon database having a plurality of elements each element associated to said each structure;

means for assigning at least one code of said each element to said each structure; and

means for retrieving said at least one code during matching to obtain said at least one matching entry.

67. The system according to claim 66, further comprising:

means for reading lexical probability files; and

means for assigning a probability value to said each element of said plurality of elements in said lexicon database.

68. The system according to claim 61, wherein each structure of said plurality of data structures is a representation of a linguistic expression.

69. The system according to claim 64, wherein said database is a thesaurus hierarchy including a root entry, said plurality of entries depending from said root entry.

70. The system according to claim 69, wherein said relative entropy value corresponding to said each entry of said plurality of entries is calculated based on an entropy value of said each entry and an entropy value of said root entry.

71. The system according to claim 66, wherein said each element in said lexicon database is a word.

72. A system for evaluating similarity among a plurality of data structures, comprising:

means for creating a plurality of entries in a database; and

means for calculating a relative entropy value corresponding to each entry of said plurality of entries.

73. The system according to claim 72, further comprising means for storing said each entry of said plurality of entries together with said corresponding relative entropy value in a compressed format.

74. The system according to claim 72, further comprising:

means for creating said plurality of entries using a tool having a graphical user interface; and

means for exporting said plurality of entries to said database.

75. The system according to claim 72, further comprising:

means for analyzing each structure of said plurality of data structures to generate at lease one substructure;

means for matching said at least one substructure of said each structure to said database to obtain at least one matching entry; and

means for generating a match value using said relative entropy value corresponding to said at least one matching entry.

76. The system according to claim 75, further comprising:

means for verifying said plurality of entries for validity;

means for extracting from a lexicon database having a plurality of elements each element associated to said each structure;

means for reading lexical probability files;

means for assigning a probability value to said each element of said plurality of elements in said lexicon database;

means for assigning at least one code of said each element to said each structure; and

means for retrieving and matching said at least one code to said database to obtain said at least one matching entry.

77. The system according to claim 76, wherein said each structure of said plurality of data structures is a representation of a linguistic expression.

78. The system according to claim 72, wherein said database is a thesaurus hierarchy including a root entry, said plurality of entries depending from said root entry.

79. The system according to claim 78, wherein said relative entropy value corresponding to said each entry of said plurality of entries is calculated based on an entropy value for said each entry and an entropy value for said root entry.

80. The system according to claim 75, wherein said each element in said lexicon database is a word.

81. A system for evaluating similarity among a plurality of data structures, comprising:

a database having a plurality of entries;

an analyzer, coupled to said database, said analyzer configured to analyze each structure of said plurality of data structure to generate at least one substructure;

a matching unit, coupled to said analyzer and said database, said matching unit configured to match said at least one substructure to at least one entry of said plurality of entries to obtain at least one matching entry; and

an entropy calculator, coupled to said matching unit and said database, configured to generate a match value using a relative entropy value corresponding to said at least one matching entry.

82. The system according to claim 81, wherein said plurality of entries are created offline using a tool having a graphical user interface and are exported to said database.

83. The system according to claim 81, wherein said entropy calculator further calculates said relative entropy value corresponding to each entry of said plurality of entries.

84. The system according to claim 83, wherein said database stores said each entry together with said corresponding relative entropy value in a compressed format.

85. The system according to claim 81, wherein said matching unit further retrieves at least one code from said at least one substructure and matches said at least one code to said at least one entry to obtain said at least one matching entry.

86. The system according to claim 81, wherein each structure of said plurality of data structures is a representation of a linguistic expression.

87. The system according to claim 81, wherein said database is a thesaurus hierarchy including a root entry, said plurality of entries depending from said root entry.

88. The system according to claim 87, wherein said relative entropy value corresponding to said each entry of said plurality of entries is calculated based on an entropy value of said each entry and an entropy value of said root entry.